图书情报工作
第66卷 第7期 2022年4月

引文持续增长型论文特征和影响因素研究*

■ 何婕君　闵超
南京大学信息管理学院　南京 210023

摘　要：[目的/意义]通过分析引文持续增长论文的表现特征和深层原因，揭示论文持续影响力的相关规律。[方法/过程]选择美国物理学会(APS)数据集作为数据源，借助BASS模型识别引文持续增长型论文，并根据1∶1原则选取引文非持续增长型论文作为对照组，比较两组论文在文献特征和引文扩散网络上的差异。[结果/结论]结果表明，引文持续增长论文具有特殊的文献特征和扩散特性：长标题文献较多，单独作者和两人合作更为常见，对前人研究依赖程度较低；此类论文引文级联网络内部有着更多的模块结构，在较短代际距离内呈现"广播式"扩散模式。

关键词：持续增长　文献特征　引文扩散　网络指标　科学影响
分类号：G251
DOI：10.13266/j.issn.0252-3116.2022.07.011

1 引言

自E. Garfield构建科学引文索引[1]以来，论文的被引情况被广泛应用于科学评价，被引频次成为衡量学术成果重要性、实用性、受关注程度以及学术影响力的重要指标。随着研究深入，学者们提出了更多的科学评价指标，但较多指标仍是基于论文引用关系计算得出。此外，引文还可以用于科学发展动态的分析，并被认为是复杂知识网络和科学动态演化的载体，利用引文分析科学知识的传递和演化正是当前研究热点之一。

在利用引文分析科学文献学术影响力时，学者们常关注被引频次在时间维度下的变化分布[2]。被引频次的时间序列构成引文曲线，也即引文轨迹，学者们对科学论文的引文轨迹进行划分，得到了不同的划分结果[3-4]。而一篇文献的引文曲线模式既受到文献自身特征的影响，还会因施引者施引行为的不同而有所区分，这些不同带来的影响不仅可以呈现在引文曲线上，也反映在引文扩散网络和引文级联结构上。

大多数文献会经历发表、兴盛到老化、死亡的过程，此为文献的引文生命周期。伴随着科学发展速度变快，有学者提出由于文献数字化访问和搜索服务的改进，旧文献被发现和利用的可能性会提高[5]，文献生命周期也得以延长。笔者观察经典引文曲线类型发现，论文发表后直到开始衰老前的特定时间段内，曲线是近乎单调增长的，单调增长的时间跨度越大，我们认为文献受到的关注和肯定程度越高，文献生命周期跨度也更长。而如果一篇论文自发表后引文曲线呈持续增长趋势，则属于引文持续增长型论文。V. Cano等[6]发现高被引的论文存在A、B两种生命周期类型——早期崛起型和稳定增长型，后者自发表后引文保持稳定增长，表现出恒定的活力。王海燕等[2]将高被引论文曲线分为持续增长分布和显峰型分布两种，其中持续增长分布曲线由于提出的理论或方法被广泛认可，从而具备持久的学术影响力。

论文的持续被引与文献的内容质量有关，也和论文的外在特征存在联系。尽管引文持续增长型论文已有部分研究提及，但是其深层原因和表现特征有待进一步探讨。笔者利用BASS模型识别出引文持续增长型论文作为实验组，同时选择引文非持续增长论文作为对照组，借鉴引文增长影响因素和引文扩散网络研究常用方法，分析两组论文指标特征，总结其对引文持续增长的影响；通过比较两组论文引文网络的各项指标，探讨引文持续增长的影响因素和知识扩散特征。

本文的研究意义在于通过科学研究促进对引文持续增长型论文的科学认识，为发现更多论文引用规律、

* 本文系国家自然科学基金项目"引文扩散理论及实证研究"(项目编号：71874077)研究成果之一。
作者简介：何婕君，硕士研究生；闵超，助理研究员，通信作者，E-mail：mc@nju.edu.cn。
收稿日期：2021-10-21　修回日期：2022-01-17　本文起止页码：110-119　本文责任编辑：易飞

何婕君，闵超. 引文持续增长型论文特征和影响因素研究[J]. 图书情报工作,2022,66(7):110 – 119.

提高单篇论文科学影响估测能力提供借鉴。

2 文献综述

2.1 学术论文的引文扩散研究

基于引文分析科学论文知识扩散研究将单篇文献视为基本单元,根据文献引用关系进行理论研究和实证总结来分析知识的产生、传播和创新演化过程。国内多位学者从论文的引用网络[7]、引用对象[8]以及引用预测[9]等不同角度进行梳理,认为主流研究主题包括:扩散测度指标构建、引文扩散网络及路径识别、扩散结构和机理等,具体如下:

2.1.1 引文扩散衡量指标

现有知识扩散衡量指标主要从速度、广度、强度和深度 4 个方面进行构建。在引文扩散速度方面,R. Rousseau 借助施引期刊数量提出文章 – 期刊扩散速度 ADS(average diffusion speed)指标[10],Y. Liu 等根据施引学科定义了学科 – 学科 ADS 指标[11]。论文知识传播速度也体现在引文累计速度上[12]。在引文扩散广度上,学者们主要研究施引论文学科覆盖程度,目前已提出了 JDF(journal diffusion factor)[13]、学科被引率[14]等指标,T. F. Frandsen 等提出可以将研究期刊扩散特征等方法应用于单篇论文知识扩散研究[15],Y. Liu 等[11]在 JDF 等基础上提出 FDB(field diffusion breadth)指标,用以表征知识扩散的广度。梁国强等[12]则引入熵的概念来测度论文影响的学科广度。对引文扩散强度的研究与广度紧密相关,如 Y. Liu 等[11]将论文的施引文献属于特定 ESI 领域的数量作为其在该 ESI 领域中的扩散强度 FDI(field diffusion intensity)。而引文扩散的深度往往反映在论文发表后产生的引文级联增长中,有关引文级联的研究有助于从微观结构上观察引文扩散[16]。

2.1.2 引文扩散网络

引文网络相关工作主要包括网络测度指标设计和网络主路径识别两个主题。在网络测度指标设计上,学者们借鉴了图论、复杂网络分析、社会网络分析等方法,构建了网络密度、连通性、中心性、小团体、特征路径长度、聚类系数等指标,并解释其在引文扩散语境下的含义。在网络主路径识别上,学者董克、张斌认为网络路径识别相关研究主要集中于引文学科知识扩散过程、路径识别、语义分析和拓扑结构拓展 4 方面[17]。近年来也出现了一些新的主题,如有学者根据文献直引、共被引、耦合关系构建三维引文网络挖掘知识主题流动特征[18],分析单篇文献在知识网络中的扩散特征[19]。

2.1.3 引文扩散模型

科学知识扩散模型的构建多借鉴其他领域的成熟模型,陈柏彤、张斌两人将现有模型划分为类传染病模型、网络生长模型和基于引文的过程模型 3 种[20],并对 SEIRZ、SIR 等经典的模型构建过程进行了梳理。近年来,学者们尝试改良现有模型或引入新模型去分析引文扩散机理,如 L. Zhang 等[21]、C. Min 等[22]进一步扩展 Bass 模型在引文扩散领域中的研究,Z. He 等借助并改进 WSB 模型为睡美人、二次爆发等非典型引文曲线的文献构建引文模型,解释其引文内在增长机制[23],S. Copiello 则引入空间自回归模型用以分析引用行为中存在的同群效应和邻里效应[24]。

2.2 论文引文增长影响因素研究

学术论文被引频次的大小及时序分布取决于众多影响因素,国内外学者从施引者认知影响、文献计量学等角度出发开展了诸多研究。以施引者认知影响为例,早期学者 G. Nigel Gilbert 提出了施引者引用著名作者的文献是为了增强自身研究可信度的可能性[25],现有研究也表明,作者引用一篇论文的动机和目的是多样的[26],这也影响着引文累积的过程。而从文献计量学角度出发的研究主要从期刊、作者、学科以及文献自身 4 个方面分析,主要采用回归分析、相关性分析以及预测模型等方法。

2.2.1 影响因素分析

现有研究从 4 个角度进行探讨。从论文作者出发,研究者们得出作者的数量[27-28]、个人竞争力[29]、名誉声望[30]、第一作者 h 指数[31]以及作者的团队合作[32]等因素与文章的引文频次存在相关关系。从学科出发,有学者以单一学科为重点研究领域,探讨学科内引文影响因素,如 P. Ball[33]和 L. Bornmann 等[34]分别从天体物理学领域与化学领域分析了引文频次影响因素。也有学者对比了不同学科领域,如李长玲等[35]研究了不同学科下论文影响力在各评价指标上的特征差异。从期刊出发,有学者论证了期刊影响因子可以促进文献被引用[36],L. Bornmann 等认为期刊影响因子可用于预测论文被引,同时也指出论文发表历史较长时 IF 预测能力减弱[37]。从文献自身角度出发的研究结合了定性和定量的分析方法,学者 M. E. Falagas 等[38]和 P. Ball[33]发现文章长度与引用呈现正相关关系,T. Liskiewicz 等[39]认为论文标题长度和目标文献的参考文献数量对引用率存在积极影响,国内姜磊等[40]则提出参考文献质量对被引的影响并不显著。

论文的被引趋势也受到上述多种因素的影响，只是不同的场景下，各因素作用强度不尽相同。S. Wuchty 等提出随着时间推移，团队合作研究成果获得更多关注的优势会不断增加[32]；J. Wang 等的研究也证实参考文献的学科多样性和差异性对长期引用有着积极影响，而对短期引用产生负面作用[41]。

2.2.2 研究方法选取

研究者们大多采用回归分析法，以论文被引频次为因变量，提取和量化影响因素指标作为自变量进行实证分析。如 L. Bornmannn 借助多级回归模型得出归一化论文影响力（NCSs）与作者数量、参考文献数量、页数和期刊影响因子关系较弱的结果[42]。也有学者采用其他的方法，如 S. Jiang 等提出了一个多维网络框架，借助社会资本理论探究研究人员的协作经验、可信度和职业生涯等因素对知识扩散产生的正向影响[43]。

2.3 述评

纵观国内外相关研究，可知两个研究领域都已经较为成熟。学者们不断加深对单篇论文的引文扩散特征、增长机理的研究，尝试构建新的指标、引入更多模型更好地解释引文扩散现象，同时从多个角度考察和探讨论文被引频次及其累计过程的影响因素，即使不同学者得出的研究结论可能并不一致。然而，目前对特定一类引文曲线的形成原因以及更细致的引文扩散特征的研究较少，也鲜有研究对引文持续增长型论文的引文增长影响因素和知识扩散特征进行探索，故笔者引入文献特征和网络特征相关指标，采用描述性统计及配对样本非参数检验方法，探讨该类论文的一般规律与特征。由于物理学最能代表基础研究领域，笔者将其作为实证领域，识别出该领域内引文呈现持续增长的科学论文，观察这部分论文与对照组文献相比所具备的文献内容特征和形式特征以及在引文网络中的异同，以期揭示引文持续增长论文知识扩散的一般特征及其影响因素。

3 数据和方法

3.1 数据

研究数据源于美国物理学会（APS）数据集，纳入了 1893－2013 年间的文献。为识别引文持续增长型论文，本研究将 Bass 模型应用到 APS 数据集上进行拟合。该模型是 F. M. Bass[44] 对创新扩散理论的创新机制和模仿机制的量化建模，是分析创新产品或技术扩散过程的常用经典模型。Bass 数学建模原理如下：

$$\frac{f(t)}{1-F(t)} = p + \frac{q}{M}[A(t)] \qquad \text{公式(1)}$$

在公式（1）中，f(t) 代表 t 时的采纳部分，$F(t)$ 为从初始到 t 时的采纳部分，$A(t)$ 为 t 时的累计采用者数量。p、q、M 分别为创新参数、模仿参数和市场总潜力（最终的采用数量）。

已有学者将该模型引入引文分析研究，其中 C. Min 等[22] 根据 Bass 模型拟合得到的参数值将科学论文划分为 4 组，其中参数 p 值和 q 值较小的论文年度引文量呈现出持续增长的特征，这部分论文学术影响力和未来潜在的科学影响都较高。笔者采用同样的方法识别引文持续增长型论文。首先对数据集进行筛选，选择文献类型为“article”，总被引次数不低于 20，且引文窗口在 10 年以上，得到 60 491 篇论文。然后采用非线性最小二乘法（NLS）检验拟合结果，提取所有非负参数值和 $R^2 \geq 0.5$ 的论文共计 8 243 篇。最后将 p 值和 q 值均小于其十分位数的 186 篇论文作为实验组论文。

为了更好地对比分析，参照已有研究[12] 按照 1:1 的比例建立对照组：对实验组中每一篇论文，选取与该论文同年发表在同一期刊上且总被引量相近的论文放入对照组，最终获取 186 篇非引文持续增长型论文。为了更好说明两组论文的区别，从实验组和对照组中各选两篇示例论文绘制引文曲线见图 1。

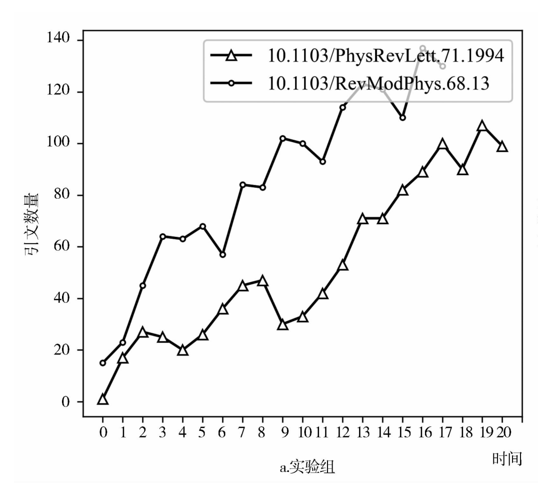
3.2 分析指标构建

为了对比两组论文被引影响因素和引文扩散特征，笔者对每篇论文均获取和计算其文献特征和引文网络两方面的特征，分别选取指标如表 1 所示。所有指标均用于对比实验组和对照组论文的特征差异，限于研究重点而未作指标内在逻辑关联的探讨。

（1）文献特征。选取论文标题长度、作者人数、论文篇幅、参考文献数量与普赖斯指数 5 个指标作为文献特征，这些指标是从常用指标以及之前一些已经证实与论文被引存在正相关关系的研究工作[46－48] 中收集的。

（2）网络特征。引文扩散研究包含科学论文施引过程的动态研究，笔者选择科学文献发表后的四代施引论文构建引文级联网络，对网络密度、平均聚集系数、最长路径长度这 3 个常用于刻画网络结构的指标进行计算来测度引文扩散的动态特征。

为了衡量文献的新颖性和创造性，笔者采纳了基于论文参考文献和施引文献来测度其科学颠覆性的 D 指数（Disruption index）指标[49]，即引用目标文献而没



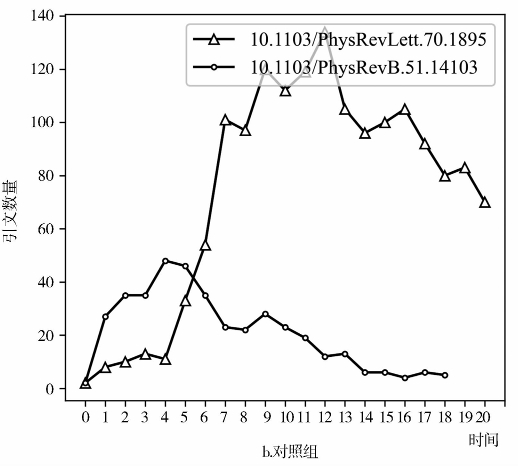


图 1　两组引文曲线示例

表 1　指标对比

特征分类	具体指标	定义及计算
文献特征	论文标题长度	论文标题所含单词个数
	论文篇幅	论文页数
	作者人数	论文合作者数量
	参考文献数量	参考文献篇数
	普赖斯指数[45]	参考文献的知识新近程度
网络特征	网络密度	网络中实际存在的边数与可容纳的边数上限的比值 $D=\frac{2\|E\|}{\|V\|(\|V\|-1)}$
	平均聚集系数	全局聚集系数用于度量整个网络的集聚性 $C=\frac{1}{\|V\|}\sum_{v_i}C(v_i)$
	最长路径长度	针对引文级联有向无环图求其关键路径的长度
	创新性	科学文献的知识颠覆性强度

有引用其参考文献的施引文献数量与同时引用目标文献及其参考文献之一的施引文献数量与目标文献及其参考文献所有施引文献之和的比例。简单来说，该指数越大，文献的颠覆性越强。D 指数计算方式如下：

$$DI_l=\frac{N_i-N_j^l}{N_i+N_j^l+N_k}\qquad \text{公式(2)}$$

其中，N_i 为引用目标文献但未引用其参考文献的施引文献数量，N_j^l 为引用了目标文献的同时至少引用了 l 篇其参考文献的施引文献数量，N_k 为没有引用目标文献，但至少引用了 1 篇目标文献参考文献的文献数量。

4　结果

笔者借助 KS－检验分析所有指标数据是否服从正态分布，结果 P 值均大于 0.05，拒绝了这一假设，因此可以采用威尔克森符号秩检验(Wilcoxon signed-rank test)比较两组数据的区别。该方法是对非参数统计中符号检验法的改进，适用于总体数据不服从正态分布且分布情况不明的情况，可被用来检验数据资料是否来自同一个总体的假设。为进一步对比和了解两组论文在具有显著差异的指标上的特征，笔者先进行统计检验，根据检验结果针对具体指标进行可视化展示和分析。

4.1　文献特征指标

表 2 与表 3 分别为两组论文文献特征指标的检验结果和描述统计。数据表明，两组论文在论文篇幅上不存在显著差异(p 值 0.144 5＞0.05)，但在标题长度、作者人数、参考文献数量、普赖斯指数指标上存在差异，图 2 即为实验组与对照组在文献特征指标上的对比示意图。

表 2　文献特征指标的威尔克森符号秩检验结果

特征	指标	检验结果
文献特征	论文标题长度	0.030 5 *
	论文篇幅	0.144 5
	作者人数	0.000 1 ***
	参考文献数量	0.028 6 *
	普赖斯指数	0.007 6 **

注：*表示 Sig. 值＜0.05，**表示 Sig. 值＜0.01，***表示 Sig. 值＜0.001

结合图 2 可以看出，两组论文标题长度主要集中在[4, 13] 区间内，引文持续增长论文的标题长度均值略高于非持续增长论文，观察数值分布可以发现，当标题中单词个数超过 8 个时，实验组论文数量较对照组更多，表明实验组中长标题论文更多；而引文非持续增长的论文标题长度更集中分布在[5,9]内。

表 3 两组论文文献特征指标的描述性统计结果

统计量	论文标题长度		论文篇幅		作者人数		参考文献数量		普赖斯指数	
	实验组	对照组	实验组	对照组	实验组	实验组	实验组	对照组	实验组	对照组
平均值	8.833	8.161	12.651	12.731	2.038	2.887	17.554	17.317	0.587	0.690
标准差	3.766	3.378	21.039	14.725	1.122	3.590	47.681	22.201	0.311	0.267
最小值	2	2	1	1	1	1	0	0	0	0
最大值	22	24	224	95	7	34	582	165	1	1
25%	6	6	3	3.25	1	1	4	6	0.377	0.500
50%	8	8	8	8.5	2	2	7	11	0.615	0.735
75%	11	10	14	15	2	3	16	19	0.833	0.912

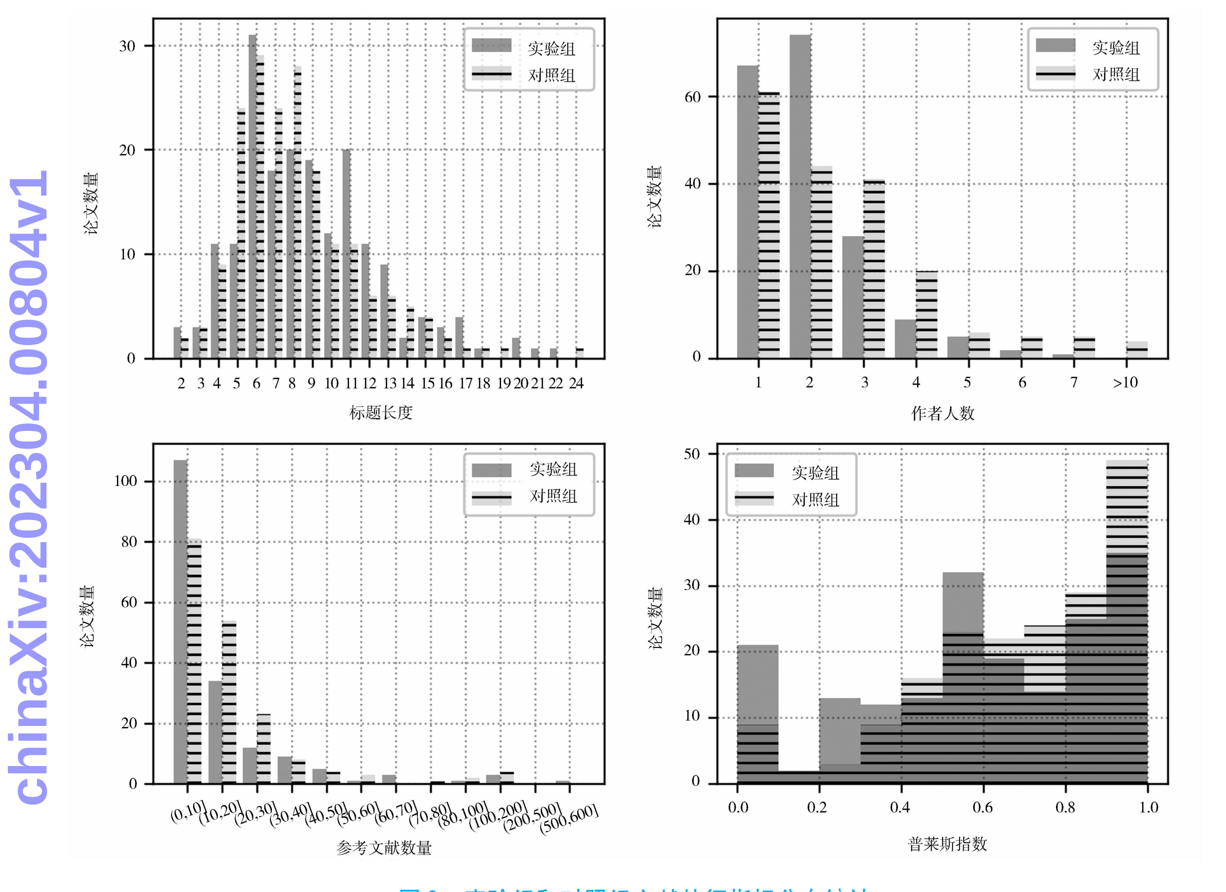


图 2 实验组和对照组文献特征指标分布统计

在论文作者人数方面，两组论文存在显著差异（p 值 0.000 1 < 0.001）。对照组论文均值为 2.038，小于实验组论文合作者数量均值 2.887，也就是说，引文非持续增长论文篇均作者人数平均比引文持续增长论文作者约多 1 人。此外，实验组论文作者数量标准差远小于对照组，分布也远比对照组更集中。

在论文参考文献数量方面，尽管实验组和对照组受极值影响，篇均参考文献数量差距非常小，但检验结果表明二者仍存在明显差异。从图 2 可以看出，两组论文均呈现参考文献数量越大论文数量越少的趋势。实验组参考文献主要分布在(0,10]、[10,20]区间内，而对照组论文在参考文献数量较多的区间内论文数量更多，分布也更为离散。

在论文普赖斯指数特征方面，两组论文也存在明显差异（p 值 0.007 6 < 0.01）。观察图 2 可以看出，实验组普赖斯指数在大部分区间内分布较为均匀，对照组则主要分布在[0.6,1]高值区间内。对照组论文普赖斯指数均值高于实验组论文均值，也就是说，引文持续增长型论文比引文非持续增长论文平均被引半衰期更长，文献老化速度更慢。

4.2 网络特征指标

表 4 和表 5 分别为两组论文扩散网络指标的检验结果和描述统计。数据表明，两组论文在网络密度上不存在显著差异（p 值 0.107 8 > 0.05），在平均聚集系

表 4　网络特征指标的威尔克森符号秩检验结果

特征	指标	检验结果
网络特征	网络密度	0.107 8
	平均聚集系数	0.000 1 ***
	最长路径长度	0.032 3 *
	Dindex(1)	0.000 0 ***

注：*表示 Sig. 值＜0.05，***表示 Sig. 值＜0.001

数、最长路径长度、D 指数指标上存在差异。图 3 即为实验组与对照组在网络特征指标上的对比示意图。

在平均聚集系数方面，两组论文存在非常显著的差异。对照组论文篇均平均聚集系数为0.149，低于实验组。对照组论文平均聚集系数高度集中于(0.1，0.2)区间内，而实验组论文平均系数分布则相对离散，这说明引文持续增长论文引文网络呈现互锁式可能性更高。

表 5　两组论文引文扩散网络特征指标的描述统计结果

统计量	网络密度		平均聚集系数		最长路径长度		Dindex1	
	实验组	对照组	实验组	对照组	实验组	对照组	实验组	对照组
平均值	0.004	0.003	0.170	0.149	20.532	18.930	0.197	0.107
标准差	0.005	0.004	0.061	0.058	8.397	7.067	0.341	0.274
最小值	0.000	0.000	0.030	0.037	5	6	－0.338	－0.242
最大值	0.028	0.040	0.435	0.461	52	40	1.0	1.0
25%	0.001	0.001	0.128	0.111	14	14	－0.023	－0.036
50%	0.002	0.001	0.164	0.142	20	18	0.049	0.002
75%	0.004	0.003	0.207	0.180	25	22	0.324	0.145

在扩散网络最长路径方面，两组论文也存在显著差异。实验组论文均值为20.532，高于对照组论文长度均值18.930，与引文非持续增长论文相比，引文持续增长型论文往往扩散路径更长，网络更为复杂。

参考 P. Azoulay[49]对 L. Wu 等[50]提出的 D 指数的描述，其含义可以解释为：如果引用某篇文章的论文也引用了该文章很多比例的参考文献，那么该文章可以被视为巩固了其所属科学领域；反之，文章可以被视为破坏其领域，对该领域或研究传统造成"破坏性"或颠覆性的改变。而那些推翻既定思维范式和开辟新科学流向的研究对科学的进步发展是至关重要的。笔者计算了 DI_1，检验结果表明，实验组和对照组论文的 D 指数存在非常显著的差异(p 值 0.000 2＜0.000 1)。从表 5 和图 3 可以看出，总体上实验组 D 指数高于对照组，在 0.3－0.8 这部分高值区间内论文数量较多，分布也更为分散。而对照组主要分布在[－0.1,0.1]之间，总体分布更为集中。

5　讨论

文献引证使得蕴含在科学工作中的研究发现可以被吸收、传承和改进，一篇具有创造性、开拓性、变革性的论文通常在其发表后会不断被学者们肯定和关注，持续激发潜在的学术研究。这类论文或是采用复杂精妙的研究方法解决某个特定研究问题，或是推翻既定理论开辟新的研究视角，又或是创造了某种新的研究工具或方法等，使得其施引量呈持续增长趋势。从科学政策视角看，我们关注科学价值评估和高质量研究早期识别都是为了激励和促进更多具有新颖性和革命性的科学研究诞生，最大化研究的投入产出结果。因此了解什么样的科学研究具有长期的、持续的影响力及其知识传播的静态分布和动态演变都是非常有价值的。

笔者以 APS 数据集为基础，根据 BASS 模型识别引文持续增长型论文，并按照 1∶1原则组配对照组，在数据确定和获取的过程考虑了时间窗口和学科差异等因素带来的差异，增强数据的可对比性。研究结果证实，引文持续增长型论文具有特殊的文献特征和扩散特性。

文献特征方面，在论文标题长度上，标题作为科学研究工作的简短概括，侧面反映研究的复杂程度和深入程度，因此长标题往往能传达更丰富的信息，当然也可能是由于长标题包含更多的关键词汇，被检索使用的可能性提高，实验组的长标题文献相对较多，在一定程度上说明长标题论文更容易激发持续引用。学者 S. Wuchty 认为，团队研究通常比个人研究能够获得更高的被引频次，并且这种优势随着时间的推移而增加[32]，

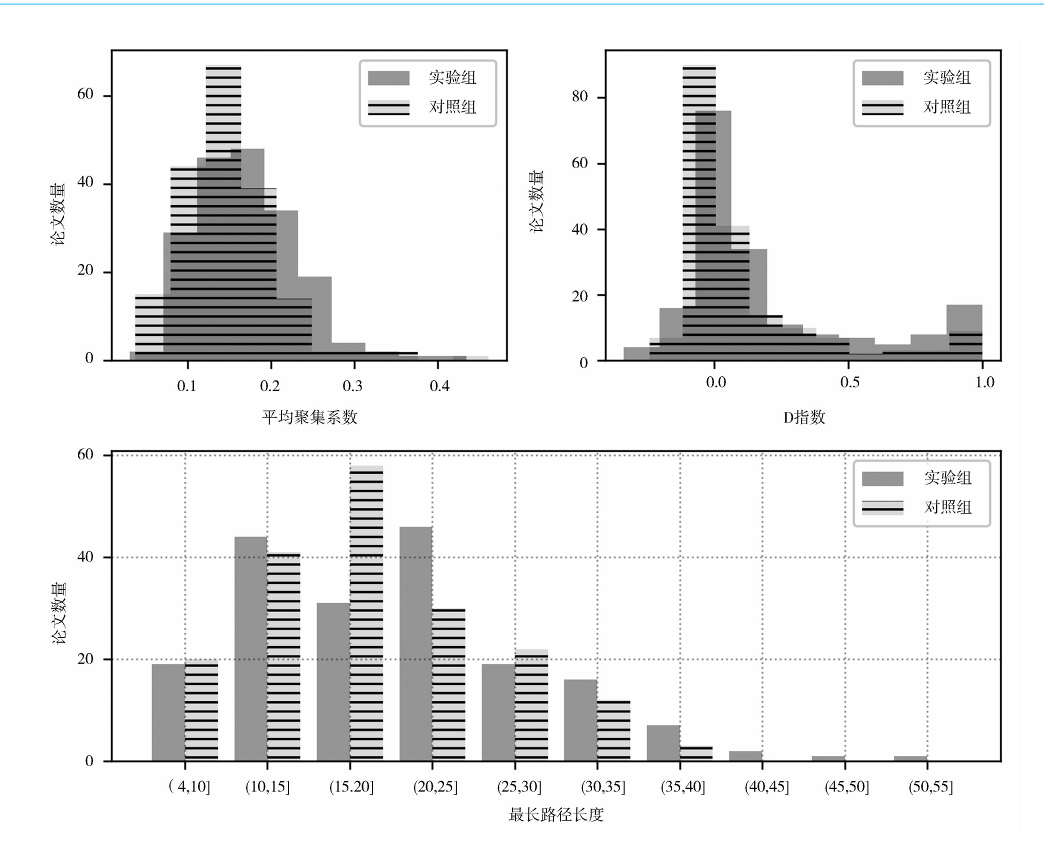


图 3 实验组和对照组网络指标分布统计

但 L. Wu 等的研究则提出单人和双人小团队比成员更多的团队更可能产出具有高影响力的论文[50]，本文的对比结果相对更支持后一种观点。引文持续增长的单篇论文作者人数往往较少也因此可以得到解释：作者人数较少的论文可能倾向于提出新的理念和思想，跳出现有的科学技术思维开启新的研究点，是潜在的高质量研究，而大团队受集体声誉和主流研究想法影响，往往在发表初期能收获大量引文，但后续引用逐渐疲软，因此没有成为引文持续增长型论文。

论文篇幅在实验组和对照组中并没有显著差异，表明文章长度对文章持续积累引文能力的区分并不明显。参考文献是科学论文的重要组成部分，反映了研究者前期工作的深度和广度，阅读吸收文献是创新工作的基础和前提，因此参考文献的数量和新颖程度也是影响文献被引的因素。两组论文对比结果说明引文持续增长型论文对领域内前人研究依赖程度并不高，更多建立在比较老旧、不太被关注的研究基础上，因而即使学者依据当下比较多、比较新的文献，甚至是立足研究热点撰写的论文，也不一定能够保持持续被引的能力。过去普遍认为随着科学发展速度加快，科学文献过时速度也会更快，然而旧文献的价值也会随着网络开放获取改进、数字化存储和搜索引擎发展而被研究人员进一步挖掘。研究表明，文献要想产生持续影响力，在重视创新性的同时也可以适当挖掘利用早期研究文献内容。

在网络特征方面，引文扩散网络反映了知识的扩散路径，也体现了知识继承、创新的过程。引文持续增长型论文的影响扩散特征也体现在其引用网络指标中。以往研究提到，平均聚集系数值越大，网络呈现互锁式可能性越高，反之，则越可能是辐射式网络[19]。引文持续增长型论文往往内部有着更多的模块结构，不同节点之间平均距离也更小，相互联系是紧密的。结合两组论文引证网络中最长路径长度的比较，笔者认为引文持续增长型文献能够在较短的代际距离内呈现“广播式”扩散，同时这种影响是深远的，存在知识的不断流动和延续。

引文持续增长型论文与引文非持续增长型论文之间的内在区别还体现在对所属学科领域的“破坏性”程度（D 指数）上，与一般科学论文相比，引证了引文持续增长型论文的研究同时引证其参考文献的可能性更低，说明该类论文成为后续研究的直接知识来源，被广泛接受和引用，即使这些研究产生的影响力比较滞后，

在发表后初期只收到了小部分的引文和关注,但之后科学价值会不断被认可。

6 结语

笔者结合文献计量部分指标与引文扩散相关理论,以物理学领域为实证领域,比较引文持续增长型论文和引文非持续增长型论文在文献特征和扩散网络特征两方面的表现和特性,分析影响论文持续吸引引用的影响因素和知识扩散模式特征。研究得出如下结论:①引文持续增长型论文往往论文标题较长,单独作者和两人合作更为常见,对前人研究依赖程度较低,更多参考比较老旧、不太被关注的研究文献,同时具备这些特征的文献相对来说更能吸引引用;②引文持续增长型论文科学颠覆性较强,其引文级联网络内部有着更多的模块结构,能够在较短代际距离内呈现"广播式"扩散模式,对未来研究产生广泛而深远的影响。

随着更多评估科学价值的方法兴起,论文持续影响力的重要性也将更加关键。引文持续增长论文的特性不仅在文献计量指标上有所预示,也反映在引用行为模式中,这既是引文持续增长论文区别于一般论文的标志,也在一定程度上揭示了其形成和扩散的成因机制。笔者对这一类引文曲线的细致研究,希望能给学者们提高自身研究成果持续影响力带来启示,为科学文献内容质量早期评估和潜在科学价值客观指标构建等相关研究提供参考,同时也服务于变革性研究的进步发展。

本研究对引文持续增长论文的分析仅是对这类论文特征的初步探索,存在一定的局限性:①缺乏对引用内容的关注,未来可以采用主题分析等方法研究目标文献在施引文献中的价值体现方式,借助回归分析等理论深入研究论文引文持续增长的具体原因和动力;②研究以物理学为实证领域,但实际上学科差异和选择的数据集均可能导致研究结果存在区别,引文持续增长型论文在不同学科和不同数据集中的特征表现仍值得进一步探索。此外,关于是否可以在发表初期就预测单篇论文的持续影响力,确定论文得以持续激发后续研究的原因有哪些以及如何测度论文的潜在持续影响力等问题,都有待今后更深入的研究。

参考文献:

[1] GARFIELD E. Citation indexes for science. a new dimension in documentation through association of ideas[J]. International journal of epidemiology, 2006, 35(5): 1123－1127.

[2] 王海燕, 马峥, 潘云涛, 等. 高被引论文与"睡美人"论文引用曲线及影响因素研究[J]. 图书情报工作, 2015, 59(16): 83－89.

[3] AVRAMESCU A. Actuality and obsolescence of scientific literature [J]. Journal of the American Society for Information Science, 1979, 30(5): 296－303.

[4] 李江,姜明利,李玥婷. 引文曲线的分析框架研究——以诺贝尔奖得主的引文曲线为例[J]. 中国图书馆学报,2014,40(2):41－49.

[5] VERSTAK A, ACHARYA A, SUZUKI H, et al. On the shoulders of giants: the growing impact of older articles[J]. arXiv preprint arXiv:1411.0275, 2014.

[6] CANO V, LIND N. Citation life cycles of ten citation classics[J]. Scientometrics, 1991, 22(2): 297－312.

[7] 陈云伟. 引文网络演化研究进展分析[J]. 情报科学, 2016, 34(8): 171－176.

[8] 马娜,张智雄,于改红. 科技论文引用对象研究综述[J]. 图书情报工作,2019,63(23):139－145.

[9] 夏琬钧,陈晓红,江艳萍. 学术论文引用预测研究进展[J]. 图书情报工作,2020,64(6):138－145.

[10] ROUSSEAU R. Robert fairthorne and the empirical power laws [J]. The journal of documentation, 2005, 61(2):194－202.

[11] LIU Y, ROUSSEAU R. Knowledge diffusion through publications and citations: a case study using ESI-fields as unit of diffusion [J]. Journal of the American Society for Information Science and Technology, 2010, 61(2): 340－351.

[12] 梁国强,侯海燕,高桐,等. 基于创新扩散理论的学术论文影响力广度研究[J]. 图书情报工作,2019,63(2):91－98.

[13] ROWLANDS I. Journal diffusion factors: a new approach to measuring research influence[J]. Aslib proceedings, 2002, 54(2):77－84.

[14] YU G, WANG M Y, YU D R. Characterizing knowledge diffusion of nanoscience & nanotechnology by citation analysis[J]. Scientometrics, 2010, 84(1):81－97.

[15] FRANDSEN T F, ROUSSEAU R, ROWLANDS I. Diffusion factors[J]. Journal of documentation, 2006, 62(1):58－72.

[16] 闵超,DING Y,李江,等. 单篇论著的引文扩散[J]. 情报学报, 2018,37(4):341－350.

[17] 董克, 张斌. 学科知识扩散网络路径识别研究进展[J]. 情报理论与实践, 2017, 40(8): 139－144.

[18] 王菲菲,王筱涵,徐硕,等. 基于三维引文关联网络的潜在知识流动探测——以基因编辑领域为例[J]. 情报学报,2021,40(2):184－193.

[19] 闵超, 张帅, 孙建军. 科学文献网络中的引文扩散——以 2011 年诺贝尔化学奖获奖论文为例[J]. 情报学报, 2020, 39(3): 259－273.

[20] 陈柏彤, 张斌. 科学知识扩散研究框架[J]. 图书情报工作, 2014(15):48－57.

[21] ZHANG L, XU K, ZHAO J. Sleeping beauties in meme diffusion [J]. Scientometrics, 2017, 112(1): 383－402.

[22] MIN C, DING Y, LI J, et al. Innovation or imitation: the diffusion of citations[J]. Journal of the Association for Information Science and Technology, 2018, 69(10): 1271 – 1282.
[23] HE Z, LEI Z, WANG D. Modeling citation dynamics of "atypical" articles[J]. Journal of the Association for Information Science and Technology, 2018, 69(9): 1148 – 1160.
[24] COPIELLO S. Peer and neighborhood effects: citation analysis using a spatial autoregressive model and pseudo-spatial data[J]. Journal of informetrics, 2019, 13(1): 238 – 254.
[25] NIGEL GILBERT G. Referencing as persuasion[J]. Social studies of science, 1977, 7(1): 113 – 122.
[26] 马凤,武夷山. 关于论文引用动机的问卷调查研究——以中国期刊研究界和情报学界为例[J]. 情报杂志,2009,28(6):9 – 14,8.
[27] BEAVER D B. Does collaborative research have greater epistemic authority? [J]. Scientometrics, 2004, 60(3): 399 – 408.
[28] 王黎明, 张啸岳, 俞立平. 论文作者数与被引频次关系的再思考[J]. 情报杂志, 2019, 38(9):166 – 170,157.
[29] SINATRA R, WANG D, DEVILLE P, et al. Quantifying the evolution of individual scientific impact[J]. Science, 2016, 354 (6312): aaf5239.
[30] PETERSEN A M, FORTUNATO S, PAN R K, et al. Reputation and impact in academic careers[J]. Proceedings of the National Academy of Sciences, 2014, 111(43): 15316 – 15321.
[31] WANG M, YU G, AN S, et al. Discovery of factors influencing citation impact based on a soft fuzzy rough set model[J]. Scientometrics, 2012, 93(3): 635 – 644.
[32] WUCHTY S, JONES B F, UZZI B. The increasing dominance of teams in production of knowledge [J]. Science, 2007, 316 (5827): 1036 – 1039.
[33] BALL P. A longer paper gathers more citations[J]. Nature, 2008, 455(7211):274 – 275.
[34] BORNMANN L, SCHIER H, MARX W, et al. What factors determine citation counts of publications in chemistry besides their quality? [J]. Journal of informetrics, 2012, 6(1): 11 – 18.
[35] 李长玲, 刘运梅, 牌艳欣. 基于学科差别的论文影响力评价指标特征分析及体系构建[J]. 情报资料工作, 2019, 40(6):23 – 29.
[36] VANCLAY J K. Factors affecting citation rates in environmental science[J]. Journal of informetrics, 2013, 7(2): 265 – 271.
[37] BORNMANN L, LEYDESDORFF L, WANG J. How to improve the prediction based on citation impact percentiles for years shortly after the publication date? [J]. Journal of informetrics, 2014, 8 (1): 175 – 180.
[38] FALAGAS M E, ZARKALI A, KARAGEORGOPOULOS D E, et al. The impact of article length on the number of future citations: a bibliometric analysis of general medicine journals[J]. Plos one, 2013, 8(2): e49476.
[39] LISKIEWICZ T, LISKIEWICZ G, PACZESNY J. Factors affecting the citations of papers in tribology journals[J]. Scientometrics, 2021, 126(4): 3321 – 3336.
[40] 姜磊, 林德明. 参考文献对论文被引频次的影响研究[J]. 科研管理, 2015, 36(1):121 – 126.
[41] WANG J, THIJS B, GLÄNZEL W. Interdisciplinarity and impact: distinct effects of variety, balance, and disparity[J]. Plos one, 2015, 10(5): e0127298.
[42] BORNMANN L. Does the normalized citation impact of universities profit from certain properties of their published documents-such as the number of authors and the impact factor of the publishing journals? a multilevel modeling approach[J]. Journal of informetrics, 2019, 13(1): 170 – 184.
[43] JIANG S, CHEN H. Examining patterns of scientific knowledge diffusion based on knowledge cyber infrastructure: a multi-dimensional network approach[J]. Scientometrics, 2019, 121(3): 1599 – 1617.
[44] BASS F M. Comments on "a new product growth for model consumer durables the bass model"[J]. Management science, 2004, 50(S): 1833 – 1840.
[45] 陈立新, 刘则渊. 引文半衰期与普赖斯指数之间的数量关系研究[J]. 图书情报知识, 2007(1):25 – 28.
[46] GNEWUCH M, WOHLRABE K. Title characteristics and citations in economics[J]. Scientometrics, 2017,110(3):1573 – 1578.
[47] 陈悦, 宋超, 周京生, 等. 文献计量学视角下的论文被引频次影响因素研究——兼评使用与被引之间关系[J]. 情报杂志, 2019,38(4):96 – 104.
[48] 葛赵青, 苗凌, 赵大良, 等. 科技期刊参考文献数量与部分引证指标的定量关系初探[J]. 编辑学报, 2015,27(5):423 – 425.
[49] AZOULAY P. Small research teams 'disrupt' science more radically than large ones[J]. Nature, 2019, 566(7744):330 – 332.
[50] WU L, WANG D, EVANS J A. Large teams develop and small teams disrupt science and technology[J]. Nature, 2019, 566 (7744):378 – 382.

作者贡献说明:
何婕君:提出研究设想,设计研究框架,收集分析数据,撰写论文;
闵超:提出研究设想,负责进行论文指导,完善研究框架。

何婕君，闵超.引文持续增长型论文特征和影响因素研究[J].图书情报工作,2022,66(7):110－119.

Research on the Features and Influencing Factors of Papers with Continuous Citation Growth

He Jiejun　Min Chao

School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/Significance] This paper reveals the relevant laws of the continuous influence of papers by analyzing the performance characteristics and deep reasons of papers with continuous citation growth. [Method/Process] This study chose the American Physical Society (APS) as the data set, used the BASS model to identify papers with continuous citation growth, and papers with non-continuous citation growth were selected as the control group according to the 1:1 principle. The two groups of papers were compared in terms of literature characteristics and citation diffusion networks to get differences. [Result/Conclusion] The results show that, papers with continuous citation growth have special literature characteristics and diffusion characteristics: long-titled papers are more, solo authors and two-person collaborations are more common, and reliance on previous research is less; and these papers have more modular structures within citation cascade networks, presenting a "broadcast" diffusion pattern over shorter intergenerational distances.

Keywords: continuous growth　literature characteristics　citation diffusion　network indicators　scientific influence